



Landslide susceptibility mapping with GIS in high mountain area of Nepal: a comparison of four methods

Pawan Gautam^{1,5} · Tetsuya Kubota² · Lok Mani Sapkota³ · Yoshinori Shinohara⁴

Received: 20 July 2020 / Accepted: 15 April 2021 / Published online: 29 April 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Landslide susceptibility mapping (LSM) assists identifying and targeting landslide preventive measures, thereby minimizing potential losses. Multiple approaches are employed for LSM in various physiographic regions; however, their applicability has differed across studies, with limited understanding on the most suitable approach for LSM in high mountain areas. Hence, we conducted LSM in the Indrawati watershed, a high mountain area of Central Nepal, employing four approaches: frequency ratio, logistic regression, artificial neural network, and support vector machine. Nine landslide causal factors (slope, aspect, elevation, geological formation, proximity to river, proximity to road, land cover, soil type, and curvature) were considered for LSM. Rainfall-induced landslides were mapped by the on-screen digitization of satellite images and field observations. The landslides were randomly split into a ratio of 80:20 for training and validating the susceptibility maps. The LSMs obtained by four methods were then validated and compared using area under curve (AUC), kappa index, and statistical inferences (sensitivity, specificity, positive predictive value, negative predictive value, and accuracy). Our study showed that Eutric Cambisols, a class of soil type, has a strong association with landslide occurrence among the 52 classes of the nine causal factors. We found that the artificial neural network approach possessed the best prediction capability (AUC value = 86.9%) among the four methods, followed by logistic regression (85.6%), support vector machine (81.2%), and frequency ratio (80.1%) approaches. However, Kappa index and other statistical inferences suggested the support vector machine approach to be the second-best method. Overall, we found that the artificial neural network yields more accurate and reliable results and hence considered as a promising approach for susceptibility mapping in high mountainous region of Hindu-Kush Himalaya. The findings of this study might be useful for landslide analysts, development planners and decision-makers in conducting LSM and development planning in high mountain regions.

Keywords Landslide susceptibility mapping · High mountain region · Frequency ratio · Logistic regression · Artificial neural network · Support vector machine

Introduction and background

Landslides are major geohazards in high mountainous regions (Ambrosi et al. 2018; Corominas et al. 2014; Petley et al. 2007). Among them, tectonically active high mountain areas, which have rough topography, and fragile and unstable geological structures, are highly prone to landslides, particularly during monsoon periods due to heavy and concentrated rainfalls (Dahal and Hasegawa 2008). Every year, landslides in high mountain areas inflict huge damages destroying villages, important infrastructure, human deaths and injuries, and causing substantial harms to national economies (Ba et al. 2018; Petley et al. 2007; Soldato et al. 2019).

Several factors contribute to landslides occurrences (Corominas et al. 2014; Lee and Pradhan 2007; Meusburger

✉ Pawan Gautam
pawan.gtm@kyudai.jp; pawan.gtm@gmail.com

¹ Laboratory of Forest Conservation and Erosion Control, Agro-Environmental Science Department, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka-shi, Fukuoka 819-0395, Japan

² Faculty of Agriculture, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka-shi, Fukuoka 819-0395, Japan

³ RECOFTC, Bangkok, Thailand

⁴ Faculty of Agriculture, University of Miyazaki, 1-1 Gakuen-Kibanadai-Nishi, Miyazaki 889-2192, Japan

⁵ Present Address: Zoological Society of London-Nepal, Gulaf Marg, Bhatbhateni, PO Box 5867, Kathmandu, Nepal

and Alewall 2008), making it difficult to accurately understand the hazard and reduce its vulnerability. While natural causes, including slope, rainfall, and seismicity are primary causes, anthropogenic factors, such as inappropriate land-use practices and infrastructure development, are also responsible for landslide occurrence (Hong et al. 2015). In such a context, preventing landslides may not be possible. However, damage to human lives and environment could be averted or reduced by identifying landslide-susceptible areas (Dai et al. 2002; Pourghasemi and Rahmati 2018) and applying appropriate mitigation measures beforehand, which is considered an effective strategy to deal with the landslide hazard (Corominas et al. 2014).

Landslide susceptibility mapping (LSM) identifies landslide-prone areas by correlating major factors responsible for landslides with past occurrence (Brenning 2005), with the assumption that similar conditions favor the occurrences of future landslides (Lee and Talib 2005; Huang and Zhao 2018). LSM helps identifying landslide-prone areas and provides information for land-use decisions to a range of users including governments, private sector, scientific communities, engineers, and planners (Akgun 2012; Ba et al. 2018; Fell et al. 2008). Furthermore, LSM is an important step in landslide risk assessment and hence beneficial to plan the mitigation measures (Michael and Samantha 2016). Reliable and accurate prediction of landslides is a difficult task because it requires high-quality data and employs a range of approaches in modelling (Yilmaz 2009). Scholars have been using several methods for LSM (Bui et al. 2016), which can be clustered in two groups: qualitative and quantitative approaches (Ayalew and Yamagishi 2005; Corominas et al. 2014; Fell et al. 2008) and the selection of method for mapping mostly depends on spatial data availability (Van Westen et al. 2008).

Qualitative approaches, such as the analytical hierarchy approach (Pourghasemi et al. 2012) and weighted linear combination (Ayalew et al. 2004), mostly depend on expert opinion and deliver the results in terms of weighted indices or relative ranks are in use (Corominas et al. 2014). In contrast, quantitative approaches, such as deterministic and statistical approaches, are based on numerical correlation of landslide distribution and causal factors (Aleotti and Chowdhury 1999). The quantitative approaches help to quantify the probability of risk in an objective manner allowing comparison and are considered important by researchers (Corominas et al. 2014). Despite their higher accuracy, deterministic or physically based models (e.g., Salciarini et al. 2017; Salvatici et al. 2018) are based on extensive experiment (e.g., Parteli et al. 2005), site-specific and data-dependent. Therefore, they are suitable for small and well-monitored landslide locations (Bui et al. 2016) and are expensive, and less practical on a regional scale (Westen and Terlien 1996). Statistical approaches analyzing the quantitative relations between past landslides and causal

factors (Fell et al. 2008) have been used for LSM. Among various statistical methods, bivariate-frequency ratio (FR) (Lee and Talib 2005; Yilmaz 2009); weight-of-evidence (Neuhauser et al. 2012; Pourghasemi et al. 2013); certainty factor (Pourghasemi et al. 2013); logistic regression (LR) (Ayalew and Yamagishi 2005; Corominas et al. 2014; Xu et al. 2013) are frequently used.

Statistical models require a large amount of data and make various statistical assumptions limiting their applicability (Pham et al. 2016). To overcome the issue, machine learning approaches including artificial neural network (ANN) (Bui et al. 2016; Chen et al. 2017; Pradhan et al. 2010; Yilmaz 2009), support vector machine (SVM) (Huang and Zhao 2018; Pham et al. 2016; Pradhan 2013; Yilmaz 2010), decision tree (Pradhan 2013), neuro-fuzzy logic (Pourghasemi et al. 2012; Vahidnia et al. 2010) are being used for LSM. However, the approaches vary in terms of accuracy and suitability of these methods under different conditions, particularly for mountainous areas, as it is not adequately studied and there still exists dispute on which approach is the best (Carrara and Pike 2008; Nefeslioglu et al. 2008). The landslide-susceptible zones varied significantly even with the minimal increment of the prediction accuracies (Tien Bui et al. 2012), and therefore identifying high-performance methods is necessary to predict the landslides risk zones accurately (Bui et al. 2016). Although, there had been several attempts to compare two or more approaches for LSM (e.g. Aditian et al. 2018; Bui et al. 2016; Moosavi and Niazi 2016; Pham et al. 2016; Pradhan and Lee 2010), extensive study to gain consensus on appropriate approaches is lacking for high mountainous areas (Ambrosi et al. 2018). Hence, the comparison of different machine learning approaches with conventional methods in high mountain area is crucial to acquire adequate background knowledge to reach some reasonable conclusions.

To address the aforementioned gap, we compared LSMs based on FR and LR as statistical approaches, and ANN and SVM as machine learning approaches in the upper part of the Indrawati Watershed in Nepal to find the best suited method for LSM in high mountainous regions. Landslides together with floods during monsoon are a major hazard in Nepal and are increasing (McAduo et al. 2018) because of the land-use changes and unplanned infrastructure development (Petley et al. 2007; Cui et al. 2019). The findings of this study would assist LSM in high mountain landscapes of Hindu-Kush Himalayan region of Nepal and other developing countries, due to the comparable context.

Study area

This study was conducted in the Indrawati watershed (85° 33' N–85° 44' N; 27° 49'–28° 07' E) of the Sindhupalchok district, located about 40 km northeast of Kathmandu, Nepal (Fig. 1).

The study area covers 364 km². Although, the altitude range of the watershed is 796–5832 m, only the area up to 4000 m is included in the study as rocky slope is a prominent feature of the area above that limit. Furthermore, the area above 4000 m in Nepal is mostly covered by snow throughout the year (Mishra et al. 2014; Shrestha and Joshi 2009) and it was difficult to identify the landslides in Google Earth based satellite imageries. The focus of this study is on rainfall-induced landslides to be specific as most of the landslide occurs in the Himalayan region is caused by rainfall (Dahal and Hasegawa 2008; Petley et al. 2007). The slope ranges from 0° to 73° with a mean value of 31.6° and 60% of the area belonging to slope ranges 20°–40°.

The average annual precipitation (between 2000 and 2015) over the four stations located in the study area was 2858 mm which ranged from 2405 to 3882 mm (<http://www.dhm.gov.np/dpc>). More than 80% of the annual precipitation was observed during the monsoon season only (i.e., from June to September), which subsequently contributed to occurrence of majority of landslides in Nepal (Petley et al. 2007).

The geology of the area is young and fragile (Regmi et al. 2014). The area is located at the fore Himalaya geomorphic unit (Precambrian age) where the main process of landform development is tectonic upliftment, weathering, erosion, and slope failure (Upreti 1999). The area is dominated by Sermanthang formation and Dhad Khola gneiss formation covering 35% and 24% area, respectively (<http://www.dmgnepal.gov.np>). Sermanthang formation mostly covers the high-altitude region of the study area which includes lithology of interbedded feldspathic schist, augen gneiss, quartzite, and biotite-feldspathic schist. However, Dhad Khola gneiss covers the lower altitudinal region including porphyroblastic gneiss, augen gneiss with a thin band of quartzite and schist, and migmatitic gneiss lithology.

Forest is the dominant land cover in the area covering 70% of the land mainly on the northern side of the watershed. This is followed by agriculture covering 23% which is practiced building small terraces. Agriculture dominates the densely populated southern side which also has rapid expansion of infrastructures, such as roads and buildings. Increasing infrastructure development such as roads often without consideration to the geomorphological context of the area is contributing factor altering the geological balances in the region, posing increased landslide risks (Petley et al. 2007; Li et al. 2017).

Methods

Landslide distribution mapping and inventory

Landslide inventory is crucial for LSM (Fell et al. 2008; Jaafari et al. 2017; Van Westen et al. 2008) and was carried

out using aerial photographs and satellite imagery (Lee and Pradhan 2007; Meten et al. 2015; Shahabi and Hashim 2015). A devastating earthquake on April 25, 2015 and its aftershocks has also hit the study area hardly triggering many landslides (Kargel et al. 2016). However, we generated pre-earthquake satellite images (December 5, 2014 and December 18, 2014), to focus on rainfall-induced landslides, from the high-resolution GeoEye satellite available on Google Earth. The location of landslide occurrences due to earthquake and rainfall is comparable as earthquake-induced landslides are often clustered along the sharper slope and depend on shaking durability, whereas the rainfall-induced landslides distributed evenly throughout the flatter slope (Gnyawali et al. 2020), where intense precipitation and higher anthropogenic pressure are observed (McAdoo et al. 2018). On-screen digitization of the satellite images in polygon format was used for the landslide inventory. The inventoried landslides were also verified by field visits (Xu et al. 2013) between November 15, 2015 and December 5, 2015, and by secondary resources such as annual reports from the relevant governmental agencies (Chalkias et al. 2014).

In total, we identified 264 rainfall-induced landslides, which cover 0.599 km² of the study area. The area of each landslide was between 757 and 108,000 m² with mean value of 2,270 m². With the help of ArcGIS 10.3, all the mapped landslides were rasterized with 30 m grid size provided by size of Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) global digital elevation model (gDEM) data.

Landslide data were divided into training landslide and validation landslide for susceptibility modelling and validating, respectively (Ba et al. 2018; Jaafari et al. 2017; Xu et al. 2013). The training dataset was employed to develop the model and the validation dataset to assess the quality of the model, in quantitative approaches of LSM (Huang and Zhao 2018). Although some studies suggested using of different ratios including 70:30 (Bui et al. 2016; Pham et al. 2016), 75:25 (Moosavi and Niazi) and even 50:50 (Pradhan 2013), Nefeslioglu et al. (2008) and Swingler (1996) suggested the use of approximately 20% of randomly selected data. Hence, we randomly selected 20% (133 pixel) of landslides as validation data and 80% (533 pixel) as training data (Fig. 1) as described in Remondo et al. (2003) (e.g., Bai et al. 2010; Jaafari et al. 2017; Nefeslioglu et al. 2008; Xu et al. 2012). After converting the landslide polygons to 30 m pixels, the landslide presence pixels were assigned to a value of 1. As landslide susceptibility modelling needs representation of no-landslide area (Xu et al. 2013), we randomly selected equal proportion of landslide absence pixels (0) from no-landslide zone of the study site (Jaafari et al. 2017; Van Den Eeckhaut et al. 2006; Xu et al. 2013) making the pixels number of training and validation dataset 1066 and 266, respectively.

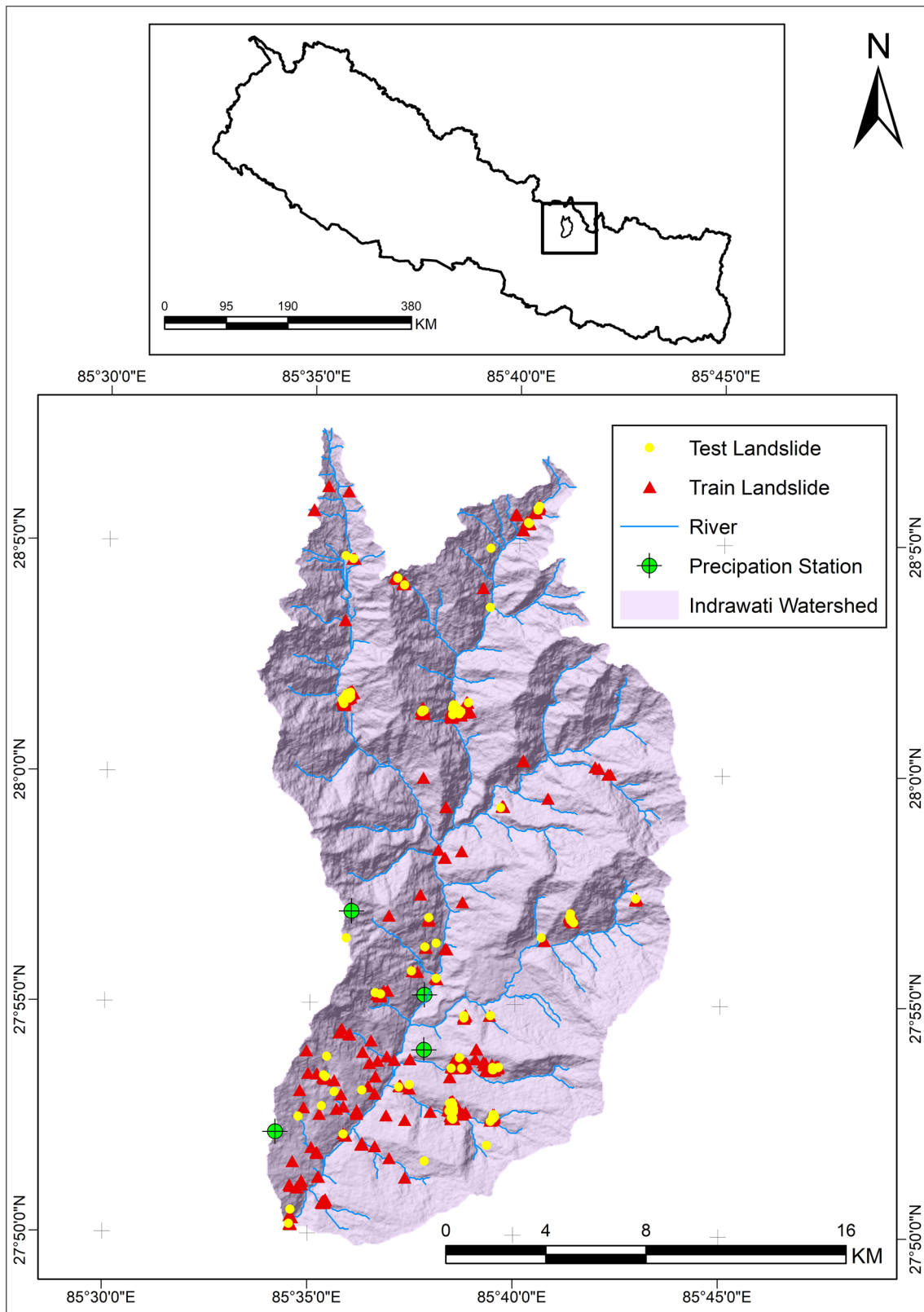


Fig. 1 Location of study along with, stream network, precipitation stations, and landslide (training and testing) locations

Landslide causal factors

Selection of landslide causal factors (independent variables contributing to landslides) is an important aspect of susceptibility mapping (Ayalew and Yamagishi 2005; Van Westen et al 2008). As suggested by Ayalew and Yamagishi (2005), the causal factors must be operational, measurable, and available for an entire study area, we used nine causal factors: slope, aspect, elevation, geological formation, land cover, proximity to river, proximity to road, soil type, and total curvature. In selecting the causal factors, the significance and source of these factors in landslides susceptibility were considered (Table 1). We used 0.7 Pearson's correlation coefficient as cut-off as suggested by Booth et al. (1994) and as employed by Bui et al. (2016) to understand the multicollinearity of conditioning factors (Table 2). A separate geo-spatial layer of each of the causal factors was developed to use in this study (Fig. 2).

Morphological factors including slope, aspect, elevation, and curvature were generated using gDEM in ArcGIS 10.3 employing surface analysis of the Spatial Analyst Tool, whereas proximity to river and road was generated by buffering the river and road features, respectively. Hydrology analysis of gDEM available on Spatial Analyst Tool in ArcGIS provided the river feature of the study area, whereas available road feature was used to buffer the road distance. The data available on ICIMOD (2013) and Dijkshoorn and Huting (2009) were used for land cover and soil type, respectively. We classified the continuous causal factor data to understand the influence of individual factors for landslide occurrences (Chen et al. 2016) in ArcGIS 10.3. Slope and elevation were classified in six discrete classes; however, curvature was classified in five classes. The classes of categorical factors including aspect (nine classes), geological formation (six classes), land cover (seven classes), and soil type (six classes) were preserved. The study area was categorized in four and two classes for proximity to river and road, respectively.

The categorical causal factors were dealt by expressing them in binary format (0 and 1) with respect to the definition of each class of causal factors whereas the non-categorical causal factors with continuous data were dealt as they are (Nefeslioglu et al. 2008). Considering five continuous variables and 28 classes of four categorical variables in binary format, total 33 independent variables were included in the LSM approaches in a data matrix where each row datum represents an individual case expressed in pixel, and columnar data show the dependent (landslide) and independent variables (causal factors). As the scales of the input variables are different, the input datum was normalized in the range of 0 and 1 as suggested by Swingler (1996) to increase the speed and accuracy of data processing, using Eq. 1 following the study of Jaafari et al. (2017) and Nefeslioglu et al. (2008):

$$X_{\text{norm}} = \frac{X_i - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}, \quad (1)$$

where X_{norm} is the normalized value of X_i input data, X_i is the input data that should be normalized, X_{max} is the maximum value of the input data, and X_{min} is the minimum value of the input data.

Analytical approaches

Frequency ratio

Using the individual relations between each causal factor and landslide occurrence in analysis (Chalkias et al. 2014; Lee and Sambath 2006), FR method allows easy and quick processing of huge amounts of data (Lee and Pradhan 2007). In this approach, the frequency ratio value (FRV), which is the ratio of the landslide ratio in each class to the area ratio (Lee and Talib 2005), was calculated for each class of each factor based on landslide densities (Corominas et al. 2014). The landslide ratio is the ratio of the landslide areas to the total area of a class. The area ratio (0.0011 in this case) is the ratio of the total landslide area to the total area of the watershed. An FRV greater than one indicates a stronger association with landslide occurrence, and vice versa; however, an FRV of one denotes the average association (Pradhan and Lee 2010). The sum of FRV of all causal factors can be defined as the landslide susceptibility index (LSI) value (Akgun 2012; Lee and Sambath 2006).

$$\text{LSI} = \sum_1^n \text{FRV}_i, \quad (2)$$

where FRV_i is the FRV of causal factor i ; and n is the number of factors (nine, in our study). All the rasters of causal factors with FRV were added to generate the combined LSI as per Eq. 2 in ArcGIS 10.3. The landslide susceptibility of each pixel is determined using LSI in the FR approach. LSI does not provide landslide susceptibility values directly, but inferences can be made considering that the higher the LSI value, the greater the area susceptible to landslides (Lee and Pradhan 2007), as the scale of landslide susceptibility is relative (Fell et al. 2008).

Logistic regression

Logistic regression, a generalized linear model (GLM) for binary response variables (Hosmer and Lemeshow 2000), is one of the commonly used approaches for landslide modeling (Brenning 2005; Bai et al. 2010; Budimir et al. 2015). The LR approach is based on analysis of a dichotomous variable for either true (landslide) or false (no-landslide), which is contributed by one or more independent variables

Table 1 Causal factors used in susceptibility mapping, their significance, data format, and sources

Causal factors	Significance in landslide	References	Data source	Data format and cell size
Slope	Describes the gravitational force (i.e., responsible for landslide occurrences) and soil strength. It is crucial conditioning and triggering factors for landslide occurrences	Corominas et al. (2014) and Porghasemi and Rahmati (2018)	ASTER gDEM ^a	Raster and 30 m
Aspect	Affects precipitation, exposure to sunlight, wind, soil moisture, and soil depth, thereby affecting the possibility of landslides	Neuhauser et al. (2012)	ASTER gDEM	Raster and 30 m
Elevation	Important topographic factors for slope instability and plays role in precipitation properties and landslide occurrences	Corominas et al. (2014) and Porghasemi and Rahmati (2018)	ASTER gDEM	Raster and 30 m
Geological formation	Strength and permeability of rock varies with lithology and formation. It plays preparatory role for landslide occurrences	Pradhan and Lee (2010) and Dahal and Hasegawa (2008)	Department of Mines and Geology, Nepal	Vector and converted to raster of 30 m
Land cover	Vegetation cover dictates soil water regulation and mechanical anchoring of soil, which is altered by changes in land cover and plays huge role for landslide occurrences	Corominas et al. (2014), Meusburger and Alewall (2008) and Petley et al. (2007)	ICIMOD 2013	Vector and converted to raster of 30 m
Proximity to river	Landslide frequency is affected by proximity to river and stream network	Corominas et al. (2014) and Meten et al. (2015)	ASTER gDEM, hydrological analysis in GIS	Vector and converted to raster of 30 m
Proximity to road	Road construction in hilly area cause adverse impacts on slope stability and contribute for increase in landslide distribution	Petley et al. (2007) and Porghasemi and Rahmati (2018)	Drawn from satellite image (270 km)	Vector and converted to raster of 30 m
Soil type	Affects the moisture retention capacity of the soil and crucial factor for landslide occurrences	Corominas et al. (2014)	Dijkshoorn and Huting (2009)	Vector and converted to raster of 30 m
Total curvature	Affects the topographical morphology and controls the soil erosion process	Corominas et al. (2014) and Gnyawali et al. (2020)	ASTER gDEM	Raster and 30 m

^aAvailable on <https://earthexplorer.usgs.gov>

Table 2 Pearson's correlation coefficient between landslide causal factors

Causal factors	Slope	Aspect	Curvature	Elevation	Landcover	Lithology	River Dis	Road Dis	Soil Type
Slope	1	–	–	–	–	–	–	–	–
Aspect	– 0.11	1	–	–	–	–	–	–	–
Curvature	0.55	– 0.01	1	–	–	–	–	–	–
Elevation	0.07	0	– 0.08	1	–	–	–	–	–
Landcover	– 0.2	– 0.02	0.02	– 0.48	1	–	–	–	–
Lithology	– 0.16	– 0.06	– 0.03	0.17	0.38	1	–	–	–
River Dis	– 0.05	– 0.03	0	0.31	– 0.07	– 0.04	1	–	–
Road Dis	0.05	– 0.04	– 0.07	0.69	– 0.35	0.3	– 0.06	1	–
Soil Type	– 0.12	– 0.06	0.06	– 0.68	0.57	0.07	– 0.2	– 0.59	1

(Menard 1995). The probability of landslide occurrence calculated based on the relation between landslide spatial distribution and causal factors (Ayalew and Yamagishi 2005; Budimir et al. 2015), is related to the logit value that contains the independent variables (i.e., causal factors; Mancini et al. 2010). One of the benefits of LR is that the independent variables can be non-linear, continuous, categorical, or a combination of both categorical and continuous (Bai et al. 2010; Menard 1995). Mathematically, the probability is described as follows:

$$P = 1/(1 + e^{-Z}), \quad (3)$$

where P is the probability of landslide occurrence and Z is the logit value that is expressed in linear form as shown:

$$Z = C_0 + C_1X_1 + C_2X_2 + \dots + C_nX_n, \quad (4)$$

where C_0 is the intercept; C_1, C_2, \dots, C_n are the coefficients of the landslide causal factors; X_1, X_2, \dots, X_n are the binary logistic regression function; and n is the number of causal factors (nine, in this study). An area with a higher probability (i.e., a lower logit value) has higher landslide susceptibility.

The linear logit model in LR (i.e., Eq. 4) represents the presence and absence of landslides on causal factors (independent variables; Bai et al. 2010). The coefficients for all causal factors were estimated using the maximum likelihood criterion (Mancini et al. 2010). The causal factor with the highest coefficient value is concluded to have the greatest effect on landslide occurrence (Budimir et al. 2015). For this purpose, a "glm" package in R 3.5.2 was used to develop a GLM (Team 2019).

Artificial neural network

Artificial neural networks consist of interconnected groups of artificial neurons that have capability to deal with complex relationships between input and output variables including landslides (Alkhasawneh et al. 2014; Bui et al. 2016; Fausett 2006; Yilmaz 2009). ANN consists of an input layer

(causal factors of landslide), one or more hidden layers and an output layer (Kavzoglu and Mather 2003).

Among the various proposed algorithms for neural networks in literature, we used multi-layered perceptron (MLP) neural net which is widely used ANN in landslide modelling (Bui et al. 2016). In MLP, each layer consists of neurons which are connected to the neurons in other layers through the weight and bias (Alkhasawneh et al. 2014). The weight between neurons was adjusted using a resilient back-propagation learning algorithm to determine the relationship between pairs of input (causal factors) and output (responses) vectors. The neurons of hidden and output layer process their inputs by multiplying each input by a corresponding weight, obtaining the sum of product, and then processing the sum using a non-linear transfer function (Aditian et al. 2018; Alkhasawneh et al. 2014; Poudyal et al. 2010).

We used a three-layer feed-forward ANN to build the relationship between input and output variables (Aditian et al. 2018) with a 9-5-1 network. The "neuralnet" package in R 3.5.2 (Günther and Fritsch 2010) was employed to develop an ANN model.

Support vector machine

Support vector machine is a machine learning approach based on structural risk minimization principle from statistical learning theory that transforms the input space into higher-dimensional feature space to find an optimal separating hyperplane (Abe 2010; Pham et al. 2019; Yilmaz 2010). The SVM is one of the most effective methods of classification with high accuracy (Yilmaz 2010) and is being used for LSM in recent times (Huang and Zhao 2018). The performance of SVM models is affected by the selection of kernel function and optimal parameters (Bui et al. 2016; Pradhan 2013; Pourghasemi and Rahmati 2018). Among the four types of kernel functions, such as linear, polynomial, sigmoid, and radial basis functions (RBF), RBF kernel is the most commonly used in landslide modelling (Bui et al.

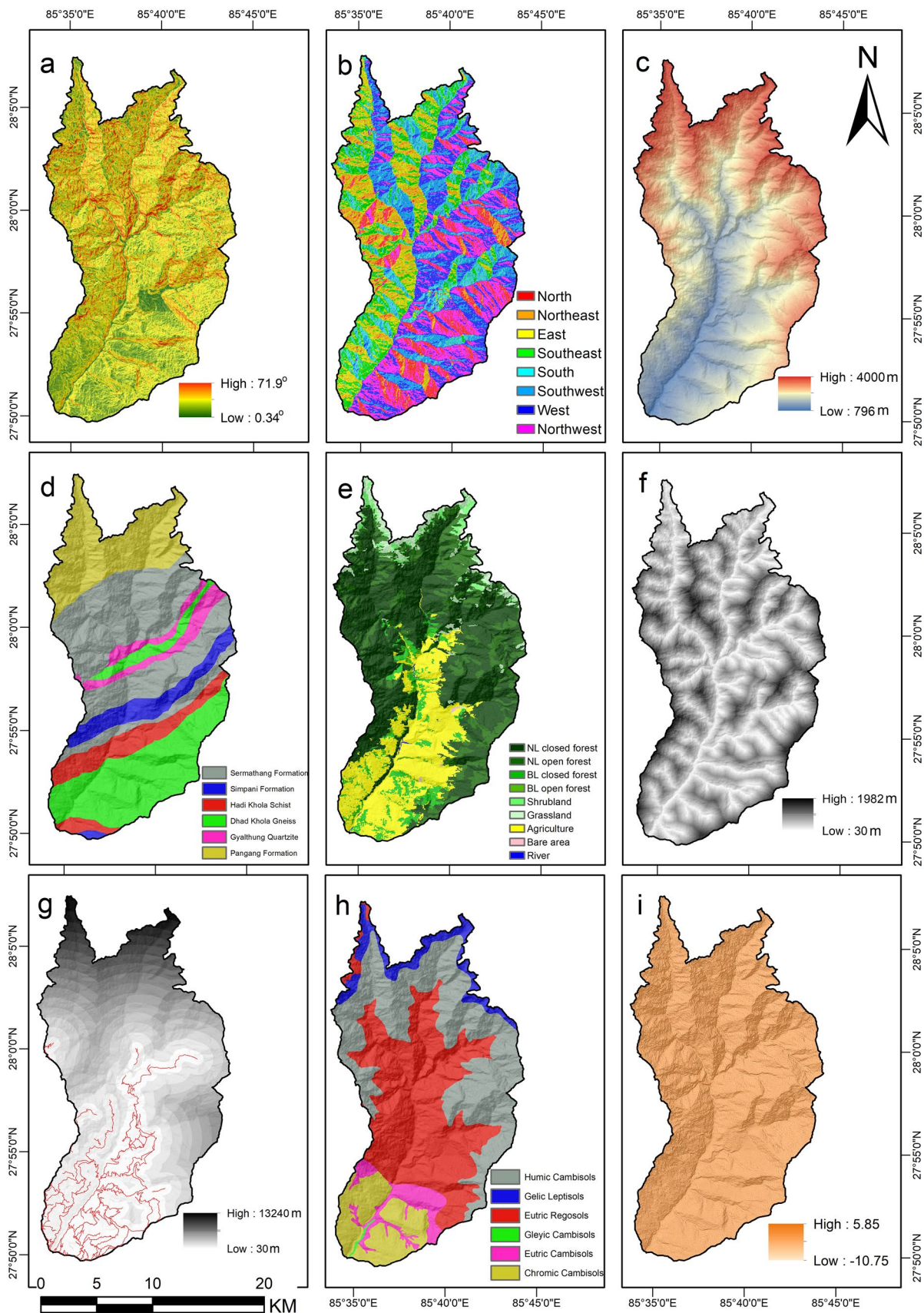


Fig. 2 Distribution of each class of the nine causal factors: **a** slope, **b** aspect, **c** elevation, **d** geological formation, **e** land cover (*NL* needle leaved, *BL* broad leaved), **f** proximity to river, **g** proximity to road (red line denotes the road feature), **h** soil type, and **i** curvature

2016; Marjanović et al. 2011; Tien Bui et al. 2012) and was also used in this study, which can be expressed as

$$K(X_i, X_j) = \exp(-\gamma \| X_i - X_j \|^2), \tag{5}$$

where γ is a kernel parameter which determines the width of the RBF and X_i, X_j are the vectors of *i*th and *j*th training landslides.

The performance of SVM using RBF is influenced by kernel width (γ) and the regularization (*C*) parameters; therefore, these parameters need to be carefully determined (Bui et al. 2016; Hong et al. 2015; Huang and Zhao 2018). The best values of the parameters were obtained using the most reliable grid search method (Kavzoglu and Colkesen 2009). A "kernlab" package in R 3.5.2 was used (Karatzoglou et al. 2004) to develop an SVM model.

LSM categories, map validation, and comparison

The obtained susceptibility values from LSM approaches were categorized in five discrete classes (very low, low, medium, high, or very high) using the natural breaks (Jenks) method in ArcGIS (Chalkias et al. 2014). Natural break method employs the natural grouping of data by putting together those with similar values and maximizing the differences between the classes.

The performances were assessed by comparing the known landslide location with landslide susceptibility maps (Xu et al. 2013). Both training and testing landslide datasets were used to assess and analyze the LSM performances (Hong et al. 2015; Pham et al. 2019). The area under curve (AUC), statistical evaluation measures, and Kappa index were used for the assessment of overall performance of the landslide models (Bui et al. 2016; Pham et al. 2016).

The susceptibility maps were overlaid with the training and validation landslides separately and AUC values were generated (Xu et al. 2012). The AUC values in the training dataset suggest the capability of the LSM approaches for mapping the landslide susceptibility (Pham et al. 2019). Whereas, the AUC values in the testing (validation) dataset indicate the prediction accuracy for future landslides (Aditian et al. 2018; Lee and Pradhan 2007; Pham et al. 2019). The AUC value lies between 0.5 and 1, where 1 indicates a perfect fit.

Five statistical evaluation measure metrics, i.e., sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV), were used to assess the performance for landslide and non-landslide classes (Hong

et al. 2015; Pham et al. 2019). These metrics are calculated employing the confusion matrices resulting from the FR, LR, ANN, and SVM models. Values for statistical indexes were calculated using following equations (Bui et al. 2016; Hong et al. 2015; Pham et al. 2016):

$$\text{Sensitivity} = \frac{TP}{FN + TP}, \tag{6}$$

$$\text{Specificity} = \frac{TN}{FP + TN}, \tag{7}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{8}$$

$$\text{PPV} = \frac{TP}{TP + FP}, \tag{9}$$

$$\text{NPV} = \frac{TN}{TN + FN}, \tag{10}$$

where TP (true positive) is the number of landslide pixels correctly classified to landslide class, TN (true negative) is the number of non-landslide points correctly classified as non-landslide. FP (false positive) is the number of landslide pixels classified to non-landslide class and FN (false negative) is the number of non-landslide points classified to landslide class (Hong et al. 2015).

Kappa index as an additional evaluator (Pham et al. 2019) was used to understand the prediction accuracy of the model, which is given as:

$$K = \frac{P_{\text{obs}} - P_{\text{exp}}}{1 - P_{\text{exp}}}, \tag{11}$$

$$P_{\text{obs}} = (TP + TN), \tag{12}$$

$$P_{\text{exp}} = (TP + FN) \times (TP + FP) + (FP + TN) \times (FN + TN), \tag{13}$$

where P_{obs} is the proportion of pixels predicted correctly as landslide or non-landslide and P_{exp} is the proportion of pixels as the agreement is expected by chance (Hoehler 2000).

Results

Frequency ratio values and landslide susceptibility

Eighteen classes among 52 classes of the nine causal factors show the positive association for landslide occurrences as per the FRV (Table 3). The slope class 30°–40° and 40°–50° had the respective values of 1.721 and 1.112 among the six

classes, indicating positive relations with landslide occurrence. Of the nine aspect classes, the southern, southwestern, and western aspects show a positive relation to landslide occurrence: the FRVs of these three aspects were 2.185, 2.095, and 2.117, respectively. In addition, the high FRV (= 2.215) of the elevation class (1300–2400) m shows a strong positive relation with landslide occurrence.

For the six geological formation classes, the Dhad Khola Gneiss and Sermanthang Formation showed the strongest associations with landslide occurrence, with FRVs of 1.538 and 1.440, respectively. Agriculture (FRV = 1.953) displayed the strongest association with landslide occurrence among the seven land cover types. FRV decreased with increasing distance from both rivers and roads. The higher FRV (= 2.507) for regions within 100 m from rivers suggests that areas closer to rivers are more susceptible to landslides.

Within road proximity classes, the higher FRV of the class below 150 m (= 1.340) suggests that areas around roads are more susceptible to landslides. Among the six classes of the soil type, Eutric Cambisols showed the strongest positive association to landslides with an FRV of 4.188, followed by Eutric Regosols (FRV = 1.701). The regions with a negative curvature value, which indicates concavity, have a stronger relation with landslide occurrence than those with a positive curvature value, which indicates convexity. The curvature classes -4.735 to -2.087 and -2.087 to -0.385 have FR values greater than one; the latter had the highest value of 1.198 (Table 3).

The southern side of the watershed along the proximity to a river is highly susceptible to the landslides as per the final susceptibility map produced by FR approach (Fig. 3a). The five landslide susceptibility categories—very low, low,

Table 3 FRV for each class of causal factor indicating the association with landslide occurrences in the study area

Class	Area pixel	Land slide pixel	FRV	Class	Area pixel	Land slide pixel	FRV
<i>Slope (degree)</i>				<i>Land cover</i>			
< 10	14,952	6	0.364	Agriculture	92,744	200	1.953 ^a
10–20	69,390	38	0.496	Bare land	582	0	0.000
20–30	126,636	81	0.579	Closed Forest	212,713	230	0.979
30–40	118,921	226	1.721 ^a	Open Forest	71,603	11	0.139
40–50	58,786	85	1.310	Water	402	0	0.000
> 50	16,217	11	0.614	Shrub land	8249	1	0.11
<i>Aspect</i>				Grassland	18,609	5	0.243
Flat	8	0	0.000	<i>Proximity to river (m)</i>			
North	38,220	5	0.118	< 100	56,009	155	2.507 ^a
Northeast	31,977	2	0.057	100–200	48,376	95	1.779
East	47,096	11	0.212	200–400	86,178	120	1.261
Southeast	52,423	27	0.467	> 400	214,339	77	0.325
South	44,354	107	2.185 ^a	<i>Proximity to road (m)</i>			
Southwest	57,939	134	2.095	≤ 150	64,199	95	1.340 ^a
West	66,766	156	2.117	> 150	340,703	352	0.936
Northwest	66,119	5	0.069	<i>Soil type</i>			
<i>Elevation (m)</i>				Humic Cambisols	149,527	17	0.103
< 1300	32,067	25	0.706	Gelic Leptisols	23,626	5	0.192
1300–1800	70,878	153	1.955	Eutric Regosols	161,892	304	1.701
1800–2400	93,226	228	2.215 ^a	Gleyic Cambisols	806	0.0005	0.001
2400–3000	90,738	21	0.210	Eutric Cambisols	22,278	103	4.188 ^a
3000–3600	80,334	13	0.147	Chromic Cambisols	46,773	18	0.349
> 3600	37,659	7	0.168	<i>Curvature</i>			
<i>Geological formation</i>				– 22.891 to 4.735	9615	8	0.762
Sermanthang Formation	140,847	224	1.440	– 4.735 to 2.087	44,393	57	1.164
Simpani	29,188	10	0.310	– 2.087 to 0.385	108,821	144	1.198 ^a
Hadi Khola Schist	35,797	20	0.506	– 0.385 to 1.317	151,118	144	0.862
Dhad Khola Gneiss	98,903	168	1.538 ^a	1.317 to 3.962	75,549	83	0.995
Gyalthung Quartzite	24,027	2	0.075	3.962 to 25.333	15,406	11	0.651
Pangang	76,140	23	0.274				

^aFRVs indicate the classes of the causal factors highly associated with landslide occurrence

medium, high, and very high—cover 18.06% (65.72 km²), 27.33% (99.47 km²), 24.93% (90.74 km²), 20.43% (74.33 km²), and 9.3% (33.66 km²) of the total area, respectively.

Model result: logistic regression

The value of McFadden's R^2 obtained from LR analysis is 0.425, indicating that the logit model shows a good fit to the dataset (McFadden, 1977). The coefficients of distance to road, slope, aspect (except 2 classes), land cover, geological formation, and soil type (except 1 class) are positive suggesting there are positive relations for landslide occurrence (Table 4) and vice versa for the causal factors with negative coefficient.

Values of the probability of landslide occurrence in each pixel, calculated using the LR coefficients of causal factors, were between 0 and 0.9993. The final susceptibility map based on the susceptibility categories is presented in Fig. 3b, depicting the high susceptibility categories lie most consistently along the proximity to a river. The five landslide susceptibility classes—very low, low, medium, high, and very high—cover 47.94% (174.49 km²), 20.15% (73.32 km²), 12.73% (46.33 km²), 9.98% (36.32 km²), and 9.19% (33.44 km²) of the total area, respectively.

Model result: artificial neural network

The training process needed 37,984 steps until all absolute partial derivatives of the error function were smaller than 0.01, the default threshold. The misclassification errors with training and testing datasets generated from the model were 0.056 and 0.162, respectively, suggesting that 5% of the pixel of training dataset and 16% pixel of the testing dataset were misclassified. The estimated weight of input variables ranged from −1104.44 to 659.30. The lowest value was for the input variables soil type₂ (a class of soil type—Gelic leptisols) to hidden layer 4 and the highest for landcover₈ (a class of landcover—Bare area) to hidden layer 5 (ESM Appendix 1). The values for probability of landslide occurrences in each grid ranged from 0 to 0.9998. On the final susceptibility map (Fig. 3c), the five landslide susceptibility classes—very low, low, medium, high, and very high—cover 72.90% (265.27 km²), 5.17% (18.8 km²), 4.46% (16.23 km²), 5.28% (19.20 km²), and 12.20% (44.4 km²) of the total area, respectively.

Model result: support vector machine

The best values of regularization C and kernel width γ , obtained from training dataset using bound constraint SVM classification were 1 and 0.023, respectively. The final susceptibility map from SVM (Fig. 3d) showed five landslide susceptibility categories—very low, low, medium, high,

and very high—generated from natural break method, cover 15.46% (56.26 km²), 50.24% (182.80 km²), 17.74% (64.56 km²), 11.68% (42.51 km²), and 4.88% (17.76 km²) of the total area, respectively.

Map validation and model comparison

Overall performances of LSM approaches were compared using AUC values which showed ANN outperforms all the approaches with an AUC value 0.987 (Fig. 4) in the training dataset suggesting the highest goodness of fit. Similarly, the AUC value in testing data was achieved 0.869 (Fig. 5) suggesting the highest prediction capability of ANN among the modelling approaches. We found LR approach is the second-best approach in this study with 0.9 and 0.856 AUC values in the training and testing datasets, respectively. The AUC value in the training dataset also indicates that there is a good correlation between the dependent and independent variables in the LR approach (Bai et al. 2010).

Similarly, statistical indices in the training dataset (Fig. 4) suggested that ANN model has the highest performance ability followed by SVM. The ANN model properly classified 96.62% (PPV) pixel of the training dataset to landslides and 95.87% (NPV) to non-landslides. The ANN model has highest sensitivity (95.90%) in the training dataset indicating that the probability of this approach to correctly classify the pixels in landslide class is 95.90%. In the case of specificity, the ANN model was able to classify 96.60% of the pixel as non-landslide in the training dataset. Further, the ANN approach correctly classified 96.25% (accuracy) of training dataset to landslide and non-landslide pixels. The kappa index value of 0.693 for ANN model suggests that there is a substantial agreement between landslide events in the model and actual landslides on the ground. As per the values of sensitivity, accuracy, NPV, and kappa index, SVM was the second-best model with the value of 86.06%, 85.93%, 86.15%, and 0.718, respectively (Fig. 4). However, in case of specificity (88.17%) and PPV (92.12%), FR model is slightly better than SVM and LR model. Simultaneously, ANN outperforms all three models (SVM, LR, and FR) in terms of sensitivity (82.27%), specificity (86.40%), accuracy (84.21%), PPV (87.22%), NPV (81.20%), and kappa index (0.693), respectively, in the testing dataset followed by SVM model (Fig. 5). However, the PPV (81.20%) value generated from FR approach is higher than both SVM (79.70%) and LR (79.70%) approaches.

Discussion

Susceptibility mapping is an important step for the landslide hazard management in mountainous regions; however, it is a challenging task particularly in high mountain areas because

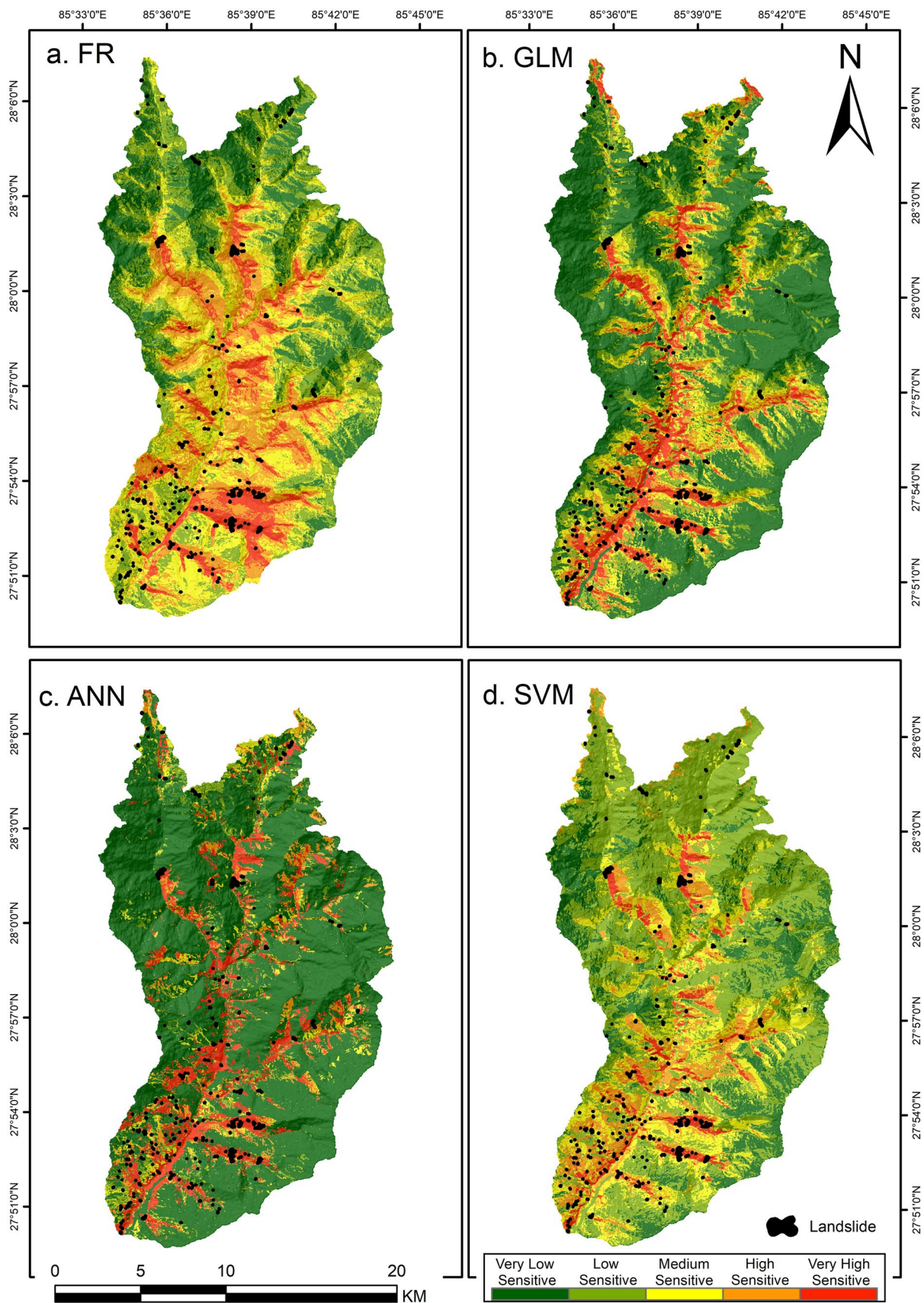


Fig. 3 Landslide susceptibility map based on: a FR, b LR, c ANN, and d SVM approaches

Table 4 LR coefficients of landslide causal factors indicating the relative significance for landslide occurrence and constant

Causal factor	Estimate	Significance	Causal Factor	Estimate	Significance
(Intercept)	-- 4.9359		Broad leaved open forest	1.3285	
Curvature	- 1.9251	**	Shrubland	0.2325	
Elevation	- 8.7061	***	Grassland	3.1243	
Distance to river	- 2.4978	***	Agriculture	2.8046	
Distance to road	8.6224	***	Bare area	4.8821	
Slope	3.3348	***	Sermanthang Frm	2.2909	***
Aspect (<45)	- 0.6754		Simpani Frm	2.8948	***
Aspect (45–90)	- 0.234		Hadikhola Schist	2.1913	***
Aspect (90–135)	1.8751	***	Dhadkhola Gneiss	2.7094	***
Aspect (135–180)	2.1438	***	Gyalthung Quartzite	1.9284	**
Aspect (180–225)	2.6094	***	Pangang Frm	NA	
Aspect (225–270)	2.776	***	Humis cambisols	0.9294	*
Aspect (270–315)	1.8314	***	Gelic leptisols	1.0558	
Aspect (315–360)	NA		Eutric regosols	0.6865	*
Needle leaved closed forest	1.9761		Gleyic cambisols	- 14.6332	
Needle leaved open forest	1.0936		Eutric cambisols	0.811	**
Broad leaved closed forest	0.4122		Chromic cambisols	NA	

Significance in *p*-value 0.001 ‘***’, *p*-value 0.01 ‘**’, *p*-value 0.05 ‘*’

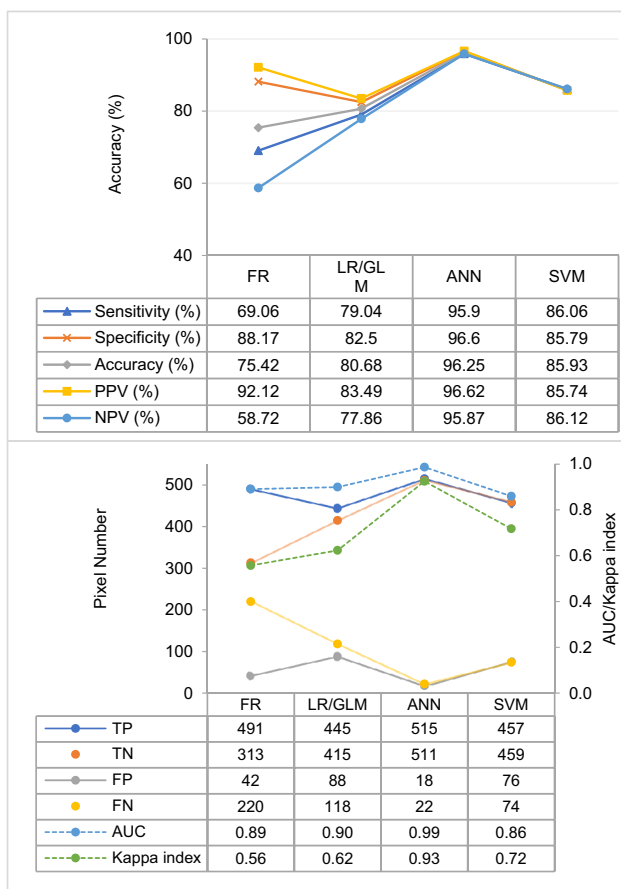


Fig. 4 Statistical indices for validation and comparison of the models using training landslide

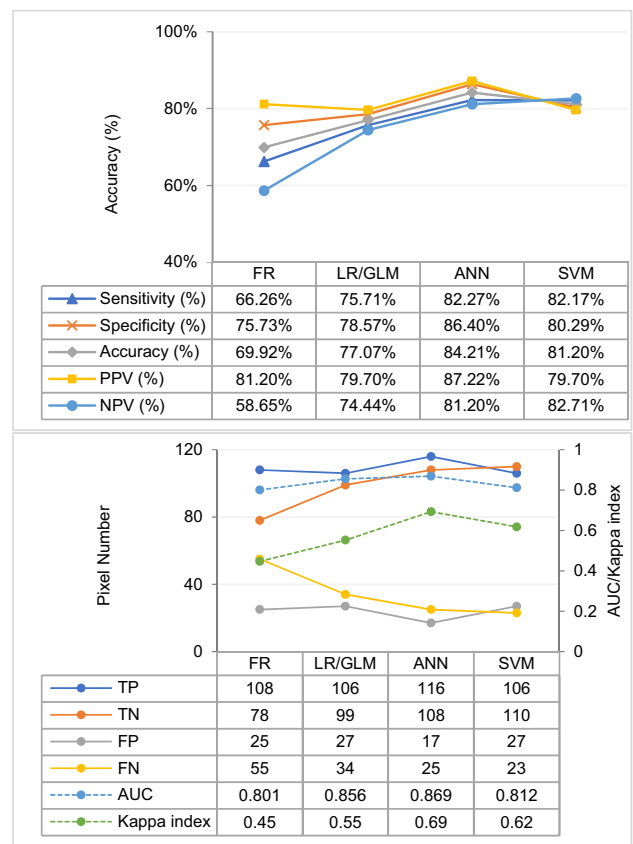


Fig. 5 Statistical indices for validation and comparison of the models using testing landslide

of difficulty in data collection due to difficult topography and cloudy weather conditions (Bui et al. 2016; Shahabi and Hashim 2015), and limited data availability. As landslide prediction can be done considering past landslide events (Tien Bui et al. 2012), we conducted a landslide inventory using satellite images available in Google Earth aided by field visits. Integration of GIS in susceptibility mapping has improved the effectiveness and ease of LSM and therefore been widely used in recent times (Huang and Zhao 2018).

Effects of causal factors on landslide occurrences

Selection of the landslide causal factors is important for LSM (Ayalew and Yamagishi 2005). Although there are no standard selection criteria, most of the LSM studies used geomorphology, topography, geology, hydrology, soil types, and land cover as causal factors (Bui et al. 2016). In the study, we used nine causal factors slope, aspect, elevation, total curvature, geological formation, land cover, proximity to river, proximity to road, and soil type for susceptibility mapping as no multicollinearity was found among them. All the causal factors used in this study area are crucial and highly relevant to the landslide occurrence in high mountain areas except aspect, which demonstrated only medium relevancy (Corominas et al. 2014).

We found that distance to road and slope has a higher positive relation to landslide occurrences through LR analysis, which also agrees with Ayalew and Yamagishi (2005) and Mancini et al. (2010). Our study suggests that the chance of landslide occurrences in high mountainous area increases with slope until the accessible slope (i.e., until the slope categories 30°–40°). This is because there are low anthropogenic activities in the higher slope as less people use the land. While, the location with high anthropogenic activities and intense rainfall favors the occurrence of rainfall-induced landslides (McAdoo et al. 2018). Our finding depicts that south facing aspects are more prone to landslides which is in line with Magliulo et al. (2008) who mapped landslides in location with hilly morphology without very steep slopes and Gnyawali et al. (2020) who mapped the landslide in hilly road. This study found that concave slopes are more susceptible to rainfall-induced landslides than convex slopes which contrast the study of Gnyawali et al. (2020). It is possibly due to concave slopes generally holding more water during rainfall, increasing the weight and contributing to slope instability (Regmi et al. 2014).

Similarly, certain elevation ranges (1300–2400 m) were more susceptible to landslides due to higher human interference. Improper land-use practices in sloppy terrain, such as rapid urbanization and conversion of forest to agricultural land (MoHA and UNDP 2009), are the reasons that those areas are susceptible to landslides. Our finding aligns with the study by Mancini et al. (2010) who found that land cover

is the important factor for landslide occurrences and the possibility of landslide occurrences is significant in arable land and urban area in a hilly terrain of Italy with maximum altitude of 1143 m, where 90% of the area is under slope angle 19°. This suggests that agricultural land and infrastructure development strategies for high mountain areas need to strongly consider the ways to mitigate landslide risks.

Areas around roads of the study site are highly susceptible to landslides in the high mountain as suggested by Ayalew and Yamagishi (2005) who found proximity to road is an important causal factor for LSM in hilly region of Japan with maximum elevation reaching to 637 m and slope ranges from 0° to 68°. The reason that the areas near the road are more susceptible to landslides (Li et al. 2017) might be due to limited technical considerations while planning and constructions of roads in the study area. As done elsewhere in Nepal and many other developing nations suffering from weak governance, suggesting weak accountability among decision-makers, and resulting limited consideration to natural ecology, hydrology and morphology in infrastructure development and land-use planning (Cui et al. 2019). The situation, along with limited technology and resources to identify areas susceptible to landslides, creates favorable conditions for landslides and the occurrences trend in Nepal is increasing (Froude and Petley 2018; McAdoo et al. 2018; Petley et al. 2007). Similar to the study by Ayalew and Yamagishi (2005), many landslides in the study area started either from the slope break or places where the topography has been altered during road construction.

Dhadkhola Gneiss and Sermanthang formations, where dominant lithology is quartzite and schist, are more susceptible to landslides in the study area. Landslide occurrences in the soil type-Eutric Cambisols and Regosols are higher. The Eutric Cambisols of the Temperate zone (part of study area is in that zone) are among the most productive soils whereas Regosols in the mountain region are relatively delicate (Driessen 1991). This might be due to the area dominated by Eutric Cambisols is more widely used for agriculture purposes.

Landslide susceptibility mapping approaches and their comparison for high mountain

Our study showed that there are several pros and cons to each model to use in LSM of high mountain areas. FR, a bivariate approach, is simple and easy, and thus, can be used to process large amounts of data and is reliable for regional-scale mapping (Lee and Sambath 2006; Yilmaz 2009). In contrast, LR, despite being a multivariate approach, is complex, produces large amounts of data, and sometimes leads to mathematical error (Lee and Pradhan 2007; Park et al. 2013; Yilmaz 2009). It assigns statistical significance value to the causal factors and produces a map taking those values into

account (Budimir et al. 2015). On the other hand, the ANN model with its superior performance has the ability to learn complex relationships between input and output variables (Pourghasemi and Rahmati 2018) that are unable to notice by either human or other computer techniques (Yilmaz 2009). The ANN is highly dependent on the experiences and prior knowledge of the designer of the neural network system (Huang and Zhao 2018). However, SVM model, a powerful machine learning tool that separates the landslides and non-landslides zone in hyperplane, is more suited if the data is relatively small and has high non-linear dimension (Yilmaz 2009; Huang and Zhao 2018). The SVM only considers the points near the boundary, called support vectors, to determine the solution of the classification problems (Pourghasemi and Rahmati 2018), whereas other approaches including LR considers each sample point for the decision plane (Huang and Zhao 2018). To identify the better suited approach for LSM in high mountainous regions, we have compared the LSMs derived from FR, LR, ANN, and SVM approaches.

The performances of LSMs were validated using different methods for comparison. All four approaches showed satisfactory accuracy in testing dataset with AUC value (0.801–0.869), kappa index (0.447–0.693), NPV (58.65–82.71%), PPV (79.70–87.22%), accuracy (69.92–84.21%), specificity (75.73–86.40%), and sensitivity (66.26–82.27%). We found ANN-derived LSM shows the higher accuracy for high mountainous areas when compared with the LSM derived from SVM, LR, and FR approaches. The finding is somehow similar to the study by Aditian et al. (2018) in the comparison of ANN, LR, and FR approaches, conducted in the hilly region of Indonesia where maximum elevation is 1207 m and 74% of the study area belongs to slopes lower than 30°. Similarly, Bui et al. (2016) and Yilmaz (2010) found that the performance of ANN with MLP is better than the SVM and LR. Likewise, Yilmaz (2009), Pradhan and Lee (2010) and Park et al. (2013) stated that LSM derived from ANN is more accurate than the LSM derived from LR and FR approaches in their respective study. Furthermore, the study by Chen et al. (2017) indicated that LSM derived by ANN has better accuracy than SVM in the study conducted in Daba mountain of China having average elevation of 1017 m and slope ranges from 0 to 80°. However, Moosavi and Niazi (2016) and Xu et al. (2012) found that SVM outperforms the LSM derived from ANN. On the other hand, Xu et al. (2012) stated that LR approach performs better than SVM and ANN approaches with better AUC value. In another study by Lee et al. (2007), LR has better prediction accuracy than ANN for LSM.

Further, LR approach shows better prediction accuracy than SVM model in LSM in high mountain area as per the AUC value (Fig. 5) which is similar to the study by Brenning (2005) and contradicts the study by Pham et al.

(2016), Moosavi and Niazi (2016) and Yilmaz (2010). Pham et al. (2016) reported that SVM outperforms LR and other machine learning approaches for LSM in the mountainous region of Uttarkhanda, India where the maximum elevation reaches 2738 m and slope 70°. However, the SVM outperforms LR as per the values of other statistical inferences including kappa index, accuracy, NPV, and PPV in this study which coincides with the study by Pham et al. (2019). Furthermore, similar to our study, SVM with RBF function outperforms the FR approach in the study by Chen et al. (2016). In contrast, Poudyal et al. (2010) found that FR has slightly better accuracy than ANN approach in their study of Panchthar district of Nepal located in Himalayan region, where the maximum elevation reaches to 2496 m. Similarly, Lee and Pradhan (2007) revealed that the accuracy of FR approach was slightly higher than that of LR, which contradicts with our study and the study by Akgun (2012) where we found LR-derived LSM have better accuracy.

The most suitable method varies across different physiographic regions. The local conditions of the causal factors and the methods of training data selection (Nefeslioglu et al. 2008; Xu et al. 2012), selection of causal factors (Pradhan and Lee 2010), mapping approaches used, and data quality (Nefeslioglu et al. 2008) are possible causes for these contrasting results and we suggest further research to confirm the causes.

Conclusion

The study contributes for systematic comparison of four methods, including conventional (FR and LR) and machine learning (ANN and SVM) techniques, for landslide susceptibility mapping in the high mountainous region of Hindu Kush Himalaya, the area in Asia with an average slope of 31.6°. The LSM based on FR, LR, ANN, and SVM was evaluated and compared using AUC and other statistical inferences including sensitivity, specificity, accuracy, PPV, NPV, and Kappa index. The ANN approaches with its highest prediction capability of 86.9% and accuracy of 84.21% with testing landslide is found to be the suitable technique for the susceptibility mapping in the high mountainous region. Hence, the result from this study concludes that the map derived from ANN approach is best suited for mitigation of rainfall-induced landslides and land-use planning in high mountainous regions. However, further studies including in simulated condition in similar regions are likely to give more insights.

This research helps landslide analysts, decision-makers, and practitioners for landslide assessment study in such areas with a number of options but limited basis of comparison to respond to their accuracy needs in LSM (Vahidnia et al. 2010). The findings of this study can also be important

to the science community while conducting rainfall triggered landslides studies in laboratories (e.g., extending the experimental work by Brown et al. (1991) and Parteli et al. (2005) including rainfall). As land uses are changing and infrastructure development is increasing in watersheds in the Hindu Kush Himalayan region and stronger consideration of landslides is necessary, the suggestion of practical methods to map landslides with high accuracy helps mitigation of landslides.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12665-021-09650-2>.

Acknowledgements The authors would like to thank the Asian Development Bank-Japanese Scholarship Program (ADB-JSP) for providing the partial financial support for the study. The authors are sincerely grateful for the valuable suggestions and inputs from Mr. Andang Soma Suriyana (PhD) and Mr. Aril Aditian (Ph.D.), the then members of Laboratory of Forest Conservation and Erosion Control, Department of Agro-environmental Sciences, Kyushu University, and Dr. Prem Prasad Paudel of Department of Forest and Soil Conservation, Nepal. The authors cordially thank Prachanda Gautam for assisting field data collection, and Mayumi Sato and Roma for the corrections on English language. The authors are indebted for the support of Mr. Shashank Sharma of Zoological Society of London-Nepal and anonymous reviewer for their constructive comments in shaping this manuscript.

Author contributions All authors have contributed for this paper.

Funding This work is partially supported by Asian Development Bank-Japanese Scholarship Program.

Availability of data and material Data will be provided upon request.

Code availability NA.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethics approval NA.

Consent to participate NA.

Consent for publication We provide our consent to publish the submitted manuscript.

References

- Abe S (2010) Feature selection and extraction. In: Support vector machines for pattern classification. Advances in pattern recognition. Springer, London, pp. 331–341
- Aditian A, Kubota T, Shinohara Y (2018) Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. *Geomorphology* 318:101–111
- Akgun A (2012) A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: a case study at İzmir, Turkey. *Landslides* 9(1):93–106
- Aleotti P, Chowdhury R (1999) Landslide hazard assessment: summary review and new perspectives. *Bull Eng Geol Environ* 58(1):21–44
- Alkhasawneh MS, Ngah UK, Tay LT, Isa NAM (2014) Determination of importance for comprehensive topographic factors on landslide hazard mapping using artificial neural network. *Environ Earth Sci* 72(3):787–799
- Ambrosi C, Strozzi T, Scapozza C, Wegmüller U (2018) Landslide hazard assessment in the Himalayas (Nepal and Bhutan) based on earth-observation data. *Eng Geol* 237:217–228
- Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* 65(1–2):15–31
- Ayalew L, Yamagishi H, Ugawa N (2004) Landslide susceptibility mapping using GIS-based weighted linear combination, the case in Tsugawa area of Agano River, Niigata Prefecture, Japan. *Landslides* 1(1):73–81
- Ba Q, Chen Y, Deng S, Yang J, Li H (2018) A comparison of slope units and grid cells as mapping units for landslide susceptibility assessment. *Earth Sci Inf* 11(3):373–388
- Bai SB, Wang J, Lu GN et al (2010) GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the three Gorges area, China. *Geomorphology* 115:23–31
- Booth GD, Niccolucci MJ, Schuster EG (1994) Identifying proxy sets in multiple linear regression: an aid to better coefficient interpretation. Research paper INT (USA)
- Brenning A (2005) Spatial prediction models for landslide hazards: review, comparison and evaluation
- Brown SR, Scholz CH, Rundle JB (1991) A simplified spring-block model of earthquakes. *Geophys Res Lett* 18(2):215–218
- Budimir MEA, Atkinson PM, Lewis HG (2015) A systematic review of landslide probability mapping using logistic regression. *Landslides* 12(3):419–436
- Bui DT, Tuan TA, Klempe H, Pradhan B, Revhaug I (2016) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* 13(2):361–378
- Carrara A, Pike R (2008) GIS technology and models for assessing landslide hazard and risk. *Geomorphology (Amsterdam)* 94(3–4)
- Chalkias C, Ferentinou M, Polykretis C (2014) GIS-based landslide susceptibility mapping on the Peloponnese Peninsula, Greece. *Geosciences* 4(3):176–190
- Chen W, Pourghasemi HR, Kornejady A, Zhang N (2017) Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. *Geoderma* 305:314–327
- Chen W, Wang J, Xie X, Hong H, Van Trung N, Bui DT, Li X (2016) Spatial prediction of landslide susceptibility using integrated frequency ratio with entropy and support vector machines by different kernel functions. *Environ Earth Sci* 75(20):1344
- Corominas J, van Westen C, Frattini P, Cascini L, Malet JP, Fotopoulou S, Catani F, Van Den Eeckhaut M, Mavrouli O, Agliardi F, Piti-lakis K (2014) Recommendations for the quantitative analysis of landslide risk. *Bull Eng Geol Environ* 73(2):209–263
- Cui Y, Cheng D, Choi CE, Jin W, Lei Y, Kargel JS (2019) The cost of rapid and haphazard urbanization: lessons learned from the Free-town landslide disaster. *Landslides* 16(6):1167–1176
- Dahal RK, Hasegawa S (2008) Representative rainfall thresholds for landslides in the Nepal Himalaya. *Geomorphology* 100(3–4):429–443
- Dai FC, Lee CF, Ngai YY (2002) Landslide risk assessment and management: an overview. *Eng Geol* 64(1):65–87
- Del Soldato M, Di Martire D, Bianchini S, Tomás R, De Vita P, Ramondini M, Calcaterra D (2019) Assessment of landslide-induced

- damage to structures: the Agnone landslide case study (southern Italy). *Bull Eng Geol Environ* 78(4):2387–2408
- Dijkshoorn JA, Huting JRM (2009) Soil and terrain database for Nepal (1.1 million) (No. 2009/01). ISRIC-World Soil Information
- Driessen PM (ed) (1991) The major soils of the world: Lecture notes on their geography, formation, properties and use. Agricultural University
- Fausett LV (2006) Fundamentals of neural networks: architectures, algorithms and applications. Pearson Education India
- Fell R, Corominas J, Bonnard C, Cascini L, Leroy E, Savage WZ (2008) On behalf of the JTC-1 Joint Technical Committee on Landslides and Engineered Slopes (2008) guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Eng Geol* 102(3–4):85–98
- Froude MJ, Petley D (2018) Global fatal landslide occurrence from 2004 to 2016. *Nat Hazard* 18:2161–2181
- Gnyawali KR, Zhang Y, Wang G, Miao L, Pradhan AMS, Adhikari BR, Xiao L (2020) Mapping the susceptibility of rainfall and earthquake triggered landslides along China-Nepal highways. *Bull Eng Geol Env* 79(2):587–601
- Günther F, Fritsch S (2010) neuralnet: training of neural networks. *R J* 2(1):30–38
- Hoehler FK (2000) Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *J Clin Epidemiol* 53(5):499–503
- Hong H, Pradhan B, Xu C, Bui DT (2015) Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. *CATENA* 133:266–281
- Hosmer DW, Lemeshow S (2000) Applied logistic regression. Wiley, New York
- Huang Y, Zhao L (2018) Review on landslide susceptibility mapping using support vector machines. *CATENA* 165:520–529
- ICIMOD (2013) Land cover of Nepal 2010. Kathmandu, Nepal: ICIMOD. <http://apps.geoportal.icimod.org/ArcGIS/rest/services/Nepal/Landcover2010/MapServer/0>. Assessed 15 Dec 2018.
- Jaafari A, Rezaeian J, Omrani MSO (2017) Spatial prediction of slope failures in support of forestry operations safety. *Croat J For Eng* 38(1):107–118
- Karatzoglou A, Smola A, Hornik K, Zeileis A (2004) kernlab—an S4 package for kernel methods in R. *J Stat Softw* 11(9):1–20
- Kavzoglu T, Colkesen I (2009) A kernel functions analysis for support vector machines for land cover classification. *Int J Appl Earth Observ Geoinf* 11(5):352–359
- Kavzoglu T, Mather PM (2003) The use of backpropagating artificial neural networks in land cover classification. *Int J Remote Sens* 24(23):4907–4938
- Kargel JS, Leonard GJ, Shugar DH, Haritashya UK, Bevington A, Fielding EJ, Fujita K, Geertsema M, Miles ES, Steiner J, Anderson E (2016) Geomorphic and geologic controls of geohazards induced by Nepal's 2015 Gorkha earthquake. *Science* 351(6269)
- Lee S, Pradhan B (2007) Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides* 4(1):33–41
- Lee S, Sambath T (2006) Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environ Geol* 50(6):847–855
- Lee S, Talib JA (2005) Probabilistic landslide susceptibility and factor effect analysis. *Environ Geol* 47(7):982–990
- Lee S, Ryu JH, Kim IS (2007) Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression, and artificial neural network models: case study of Youngin, Korea. *Landslides* 4(4):327–338
- Li G, Lei Y, Yao H, Wu S, Ge J (2017) The influence of land urbanization on landslides: an empirical estimation based on Chinese provincial panel data. *Sci Total Environ* 595:681–690
- Magliulo P, Di Lisio A, Russo F, Zelano A (2008) Geomorphology and landslide susceptibility assessment using GIS and bivariate statistics: a case study in southern Italy. *Nat Hazards* 47(3):411–435
- Mancini F, Ceppi C, Ritrovato G (2010) GIS and statistical analysis for landslide susceptibility mapping in the Daunia area (Italy)
- Marjanović M, Kovačević M, Bajat B, Voženílek V (2011) Landslide susceptibility assessment using SVM machine learning algorithm. *Eng Geol* 123(3):225–234
- McAdoo BG, Quak M, Gnyawali KR, Adhikari BR, Devkota S, Rajbhandari PL, Sudmeier-Rieux K (2018) Roads and landslides in Nepal: how development affects environmental risk. *Nat Hazard* 18(12):3203–3210
- McFadden D (1977) Quantitative methods for analyzing travel behavior of individuals: some recent developments. Institute of Transportation Studies, University of California, Berkeley
- Menard S (1995) An introduction to logistic regression diagnostics. *Appl Logist Regress Anal* 58–79
- Meten M, Bhandary NP, Yatabe R (2015) GIS-based frequency ratio and logistic regression modelling for landslide susceptibility mapping of Debre Sina area in central Ethiopia. *J Mt Sci* 12(6):1355–1372
- Meusburger K, Alewell C (2008) Impacts of anthropogenic and environmental factors on the occurrence of shallow landslides in an alpine catchment (Urseren Valley, Switzerland). *Nat Hazard* 8:509–520
- Michael EA, Samanta S (2016) Landslide vulnerability mapping (LVM) using weighted linear combination (WLC) model through remote sensing and GIS techniques. *Model Earth Syst Environ* 2(2):88
- Mishra B, Babel MS, Tripathi NK (2014) Analysis of climatic variability and snow cover in the Kaligandaki River Basin, Himalaya, Nepal. *Theor Appl Climatol* 116(3–4):681–694
- MoHA DN, UNDP O (2009) Nepal disaster report: the hazardscape and vulnerability
- Moosavi V, Niazi Y (2016) Development of hybrid wavelet packet-statistical models (WP-SM) for landslide susceptibility mapping. *Landslides* 13(1):97–114
- Nefeslioglu HA, Gokceoglu C, Sonmez H (2008) An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. *Eng Geol* 97(3–4):171–191
- Neuhäuser B, Damm B, Terhorst B (2012) GIS-based assessment of landslide susceptibility on the base of the weights-of-evidence model. *Landslides* 9(4):511–528
- Park S, Choi C, Kim B, Kim J (2013) Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. *Environ Earth Sci* 68(5):1443–1464
- Parteli EJ, Gomes MAF, Brito VP (2005) Nontrivial temporal scaling in a Galilean stick-slip dynamics. *Phys Rev E* 71(3):036137
- Petley DN, Hearn GJ, Hart A, Rosser NJ, Dunning SA, Owen K, Mitchell WA (2007) Trends in landslide occurrence in Nepal. *Nat Hazards* 43(1):23–44
- Pham BT, Jaafari A, Prakash I, Bui DT (2019) A novel hybrid intelligent model of support vector machines and the MultiBoost ensemble for landslide susceptibility modeling. *Bull Eng Geol Environ* 78(4):2865–2886
- Pham BT, Pradhan B, Bui DT, Prakash I, Dholakia MB (2016) A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India). *Environ Model Softw* 84:240–250
- Poudyal CP, Chang C, Oh HJ, Lee S (2010) Landslide susceptibility maps comparing frequency ratio and artificial neural networks: a case study from the Nepal Himalaya. *Environ Earth Sci* 61(5):1049–1064

- Pourghasemi HR, Rahmati O (2018) Prediction of the landslide susceptibility: Which algorithm, which precision? *CATENA* 162:177–192
- Pourghasemi HR, Pradhan B, Gokceoglu C (2012) Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. *Nat Hazards* 63(2):965–996
- Pourghasemi HR, Pradhan B, Gokceoglu C, Mohammadi M, Moradi HR (2013) Application of weights-of-evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran. *Arab J Geosci* 6(7):2351–2365
- Pradhan B (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput Geosci* 51:350–365
- Pradhan B, Lee S (2010) Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ Model Softw* 25(6):747–759
- Pradhan B, Lee S, Buchroithner MF (2010) A GIS-based back-propagation neural network model and its cross-application and validation for landslide susceptibility analyses. *Comput Environ Urban Syst* 34(3):216–235
- Regmi AD, Yoshida K, Pourghasemi HR, Dhital MR, Pradhan B (2014) Landslide susceptibility mapping along Bhalubang—Shiwapur area of mid-Western Nepal using frequency ratio and conditional probability models. *J Mt Sci* 11(5):1266–1285
- Remondo J, González A, De Terán JRD, Cendrero A, Fabbri A, Chung CJF (2003) Validation of landslide susceptibility maps; examples and applications from a case study in Northern Spain. *Nat Hazards* 30(3):437–449
- Salciarini D, Fanelli G, Tamagnini C (2017) A probabilistic model for rainfall—induced shallow landslide prediction at the regional scale. *Landslides* 14(5):1731–1746
- Salvatici T, Tofani V, Rossi G, D’Ambrosio M, Stefanelli CT, Masi EB, Catani F (2018) Application of a physically based model to forecast shallow landslides at a regional scale. *Nat Hazard* 18(7):1919–1935
- Shahabi H, Hashim M (2015) Landslide susceptibility mapping using GIS-based statistical models and Remote sensing data in tropical environment. *Sci Rep* 5(1):1–15
- Shrestha AB, Joshi SP (2009) Snow cover and glacier change study in Nepalese Himalaya using remote sensing and geographic information system. *J Hydrol Meteorol* 6(1):26–36
- Swingler K (1996) Applying neural networks: a practical guide. Morgan Kaufmann
- Team RC (2019). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Tien Bui D, Pradhan B, Lofman O, Revhaug I (2012) Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and Naive Bayes Models. *Math Probl Eng* 2012
- Upreti BN (1999) An overview of the stratigraphy and tectonics of the Nepal Himalaya. *J Asian Earth Sci* 17(5–6):577–606
- Vahidnia MH, Alesheikh AA, Alimohammadi A, Hosseinali F (2010) A GIS-based neuro-fuzzy procedure for integrating knowledge and data in landslide susceptibility mapping. *Comput Geosci* 36(9):1101–1114
- Van Den Eeckhaut M, Vanwalleghem T, Poesen J, Govers G, Verstraeten G, Vandekerckhove L (2006) Prediction of landslide susceptibility using rare events logistic regression: a case-study in the Flemish Ardennes (Belgium). *Geomorphology* 76(3–4):392–410
- Van Westen CJ, Castellanos E, Kuriakose SL (2008) Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview. *Eng Geol* 102(3–4):112–131
- Westen CV, Terlien MJT (1996) An approach towards deterministic landslide hazard analysis in GIS. A case study from Manizales (Colombia). *Earth Surf Process Landf* 21(9):853–868
- Xu C, Xu X, Dai F, Saraf AK (2012) Comparison of different models for susceptibility mapping of earthquake triggered landslides related with the 2008 Wenchuan earthquake in China. *Comput Geosci* 46:317–329
- Xu C, Xu X, Dai F, Wu Z, He H, Shi F, Xu S (2013) Application of an incomplete landslide inventory, logistic regression model and its validation for landslide susceptibility mapping related to the May 12, 2008 Wenchuan earthquake of China. *Nat Hazards* 68(2):883–900
- Yilmaz I (2009) Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat—Turkey). *Comput Geosci* 35(6):1125–1138
- Yilmaz I (2010) Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environ Earth Sci* 61(4):821–836

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.